# 基于改进YOLOX的自然环境中火龙果检测方法

商枫楠[1,2,3]，周学成[1,2,3*]，梁英凯[1,2,3]，肖明玮[1,2,3]，
陈　桥[1,2,3]，罗陈迪[1,2,3]

（1. 华南农业大学 工程学院，广东广州 510642；2. 广东省农业人工智能重点实验室，广东广州 510642；3. 南方农业机械与装备关键技术教育部重点实验室，广东广州 510642）

**摘　要**：自然环境下果实的精准检测是火龙果采摘机器人执行采摘作业的先决条件。为提高自然环境下果实识别的精确性、鲁棒性和检测效率，本研究对YOLOX（You Only Look Once X）网络进行改进，提出了一种含有注意力模块的目标检测方法。为便于在嵌入式设备上部署，本方法以YOLOX-Nano网络为基准，将卷积注意力模块（Convolutional Block Attention Module，CBAM）添加到YOLOX-Nano的主干特征提取网络中，通过为主干网络提取到不同尺度的特征层分配权重系数来学习不同通道间特征的相关性，加强网络深层信息的传递，降低自然环境背景下对火龙果识别的干扰。对该方法进行性能评估和对比试验，经过训练后，该火龙果目标检测网络在测试集的$AP_{0.5}$值为98.9%，$AP_{0.5:0.95}$的值为72.4%。在相同试验条件下对比其它YOLO网络模型，该方法平均检测精度分别超越YOLOv3、YOLOv4-Tiny和YOLOv5-S模型26.2%、9.8%和7.9%。最后对不同分辨率的火龙果果园自然环境下采集的视频进行实时测试。试验结果表明，本研究提出的改进YOLOX-Nano目标检测方法，每帧平均检测时间为21.72 ms，$F_1$值为0.99，模型大小仅3.76 MB，检测速度、检测精度和模型大小满足自然环境下火龙果采摘的技术要求。

**关键词**：水果采摘；自然环境；火龙果；目标检测；YOLOX；注意力机制；深度学习

**中图分类号**：TP18　　**文献标志码**：A　　**文章编号**：SA202207001

引用格式：商枫楠, 周学成, 梁英凯, 肖明玮, 陈桥, 罗陈迪. 基于改进YOLOX的自然环境中火龙果检测方法[J]. 智慧农业(中英文), 2022, 4(3): 120-131.

SHANG Fengnan, ZHOU Xuecheng, LIANG Yingkai, XIAO Mingwei, CHEN Qiao, LUO Chendi. Detection method for dragon fruit in natural environment based on improved YOLOX[J]. Smart Agriculture, 2022, 4(3): 120-131. (in Chinese with English abstract)

## 1　引　言

火龙果因具有很高的营养、药用、保健和经济价值，其种植产业近年来在中国发展迅猛。2018年，中国已经成为世界第二大火龙果生产国，种植面积超过5万公顷，产量超过100万吨[1,2]。然而，随着国内劳动力短缺及人工费用的日益增高，火龙果生产成本变得越来越高。据统计，中国水果采收用工成本占生产总成本的35%到40%，机械采收率仅为2.33%[3]。因此，实现火龙果采收作业的机械化和自动化对提高火龙果生产的技术水平、促进火龙果产业的持续发

收稿日期：2022-06-30

基金项目：国家重点研发计划项目（2017YFD0700602）

作者简介：商枫楠（1994—），男，硕士研究生，研究方向为机器视觉与图像分析。E-mail：shangfengnan@163.com

*通信作者：周学成（1968—），男，博士，教授。研究方向为机器视觉与智能检测。E-mail：zxcem@scau.edu.cn

展具有重要的意义。而如何实现自然环境下果园中火龙果果实的精准识别和快速检测，是火龙果采收作业自动化必须解决的首要问题。

基于深度学习的神经网络目标检测方法在农产品检测领域中得到广泛应用[4]。Jordi等[5]使用Mask R-CNN实例分割网络实现苹果的检测和分割。刘志军等[6]将轻量化改进YOLOv5网络部署在无人机上实现了树上苹果的检测。Ji等[7]对MobileNetV3-Small网络进行改进，利用鲸鱼算法优化模型比例系数，优化后的模型识别多类苹果的平均精度达到94.43%。穆龙涛等[8]对Faster R-CNN中主干特征提取网络进行改进后，其对猕猴桃平均识别精度达到94.75%。在荔枝检测研究中，陈燕等[9]改进YOLOv3网络检测荔枝串，平均精度均值达到94.3%。彭红星等[10]改进SSD（Single Shot MultiBox Detector）网络对无人机拍摄的荔枝图像进行小目标检测，改进后的网络平均精确率在55.79%，比原始方法提高了约30%。

在芒果检测研究中，Roy等[11]改进YOLOv4网络，用于复杂果园环境下对芒果的检测，其平均检测精度达到96.2%。Xu等[12]提出一种改进的轻量化YOLOv3网络，实现了对自然环境下绿色芒果的检测。

在番茄检测研究中，Xu等[13]提出一种改进的Mask R-CNN网络的方法，番茄识别准确率提高了11.53%。周云成等[14]对VGGNet网络进行结构优化，对番茄果、花和茎进行检测，检测平均精度分别为81.64%、84.48%和53.94%。Zheng等[15]改进YOLOv4网络提高自然环境中番茄的检测准确率，其平均检测准确率为94.44%。赵春江等[16]利用改进的YOLOv3网络实现番茄花朵不同花期的辨识检测，并在番茄授粉机器人上进行了实际验证。

目前，针对自然环境中火龙果的目标识别与检测方法的相关研究报道较为少见，尚缺乏相对成熟的技术方法可以应用。Li等[17]和王金鹏等[18]分别改进YOLOv3和YOLOv4网络，提出轻量级卷积网络MobileNet-YOLO和YOLOv4-Tiny，火龙果检测的平均精度分别达到97%和96.48%。然而，上述对火龙果的检测研究中，其模型训练过程较为繁琐，需要手动构建锚框（Anchor）。锚框对检测结果存在影响，并且对于远处的小目标火龙果，上述方法存在漏检测的情况，对小目标果实检测效果不明显。另外，上述方法受环境和光照的影响较大，鲁棒性不高。

为了探索自然环境下成熟火龙果的精准识别和快速检测方法，针对现有研究方法存在的技术问题，本研究通过引入注意力机制，提出了一种基于改进YOLOX网络的火龙果目标检测算法，该方法在复杂背景下具有较好的鲁棒性，适合自然环境下火龙果的目标检测，其结果可为其他复杂环境下水果的快速识别提供一定的参考。

## 2 材料与方法

### 2.1 试验数据获取

由于缺少可直接使用的火龙果数据集，本研究在广州市增城区和从化区火龙果种植园拍摄火龙果图像，时间为2021年7月17日（晴天）和2021年10月17日（遮阴天）。拍摄设备为CANON EOS M100相机，成像距离为50～150 cm，自动曝光，图像分辨率像素为2400×1344，以*.JPG格式存储。共采集原始图像1976张，包括在晴天顺光、晴天逆光、阴天视角下的整株多果和任意单果图像（图1）。为降低重复图片数量以及无果实图片对模型训练的干扰，使用人工筛选的方法对采集到的原始图像进行数据清洗，清洗后得到总计1744张包含火龙果的原始图像。

### 2.2 数据集准备

使用开源LabelImg工具手工为图像中的火龙果添加标签，并参考Pascal VOC制作数据集，标签名称为Dragon_fruit。标签原则为在视野范围内，根据火龙果表面颜色信息进行人工标记。具体为：对表皮红色像素面积超过90%的火龙果进行标记，不考虑绿色果实；有遮挡的火龙果，


(a)晴天逆光


(b)晴天顺光


(c)阴天

图1 不同拍摄环境下火龙果原始图像

Fig. 1 Original images of dragon fruit in different environments

人工估计被遮挡区域进行标记；视野内存在较远处的果实，若像素面积过小（≤20×20）则不进行标记。

## 2.3 数据集扩充

数据集扩充可以有效地扩展样本的多样性，保证模型在不同环境下具有较高的鲁棒性。在本研究中，使用了旋转、翻转、添加噪声、模糊等操作将数据集扩充至5232张。

# 3 识别网络模型构建

YOLO（You Only Looks Once）算法是一阶段目标检测算法的代表之一，由于实时检测速度快，检测精度高，在农业领域的目标检测和识别中得到广泛应用。YOLOv1～YOLOv5均是基于锚框的探测器，需要手工预先设定锚框大小的分配规则。

## 3.1 YOLOX网络模型

由于本研究是针对采摘机器人在田间自然环境下的实时检测，因此研究的目标检测网络需要考虑的主要因素有模型大小是否易于在嵌入式设备上部署，在复杂光照及背景条件下检测精度和推理速度。YOLOX是Ge等[19]提出的单阶段目标检测算法，它将目标区域预测和目标类别预测集成到一个单一的神经网络模型中，融合了YOLO系列网络的优点，创新地添加了收敛速度更快、精度更高的解耦头、并引入无锚框（Anchor free）方法和SimOTA（标签分配策略）动态正样本匹配等[20-22]，实现了高精度的快速目标检测和识别。同时，YOLOX有多种基准的型号用于不同的应用场景，例如YOLOX-X、YOLOX-L和YOLOX-Darknet53等标准型号和采用深度可分离卷积构建的轻量化型号。

### 3.1.1 网络结构

YOLOX使用CSPDarkNet作为主干特征提取网络。输入图像在主干特征提取网络中进行特征提取，提取到的特征可以被称作特征层，作为输入图像的特征集合。在CSPDarkNet中，获取到三个特征层作为有效特征层，它们位于CSPDarkNet的不同位置，为下一步的网络建设做准备。

特征金字塔网络可以被称作YOLOX的加强特征提取网络，在CSPDarkNet中获取到的三个有效特征层会在这一部分进行特征融合，以获得不同尺度的特征信息。YOLOX中同样使用了在YOLOv4中用到的路径聚合网络（Path Aggregation Network，PANet）结构[23]。PANet在自上向下的特征金字塔网络（Feature Pyramid Network，FPN）结构中增加了自底向上的路径聚合模块，以更好地传达语义和位置信息。PANet能够缩短信息路径，增强特征金字塔，对于YOLOX网络实现多尺度检测，尤其是对小目标的检测非常重要。

YOLOX的解耦头（Decoupled head）与以往的YOLO不同，它包含一个1×1的卷积层以调整通道数量，然后是两个并行分支，每个分支上包含两个卷积层，两条分支分别用于分类和回归任务，计算重叠度（Intersection over Union，IoU）分支被添加在回归分支上。

### 3.1.2 损失函数

本研究使用的YOLOX的损失函数由三部分

组成，总损失见公式（1），

$$L = L_{Reg} + L_{Cls} + L_{Obj} \quad (1)$$

其中，$L$为总损失，$L_{Reg}$为预测框与真实框之间的位置误差的损失，$L_{Cls}$为判断对象类别概率的损失，$L_{Obj}$为对象置信度的损失；*Cls*和*Obj*分别表示对象的类别概率和置信度，由二值交叉熵损失函数计算得到。公式（2）为IoU损失函数表达式。

$$Loss_{IoU} = 1 - IoU = 1 - \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

其中，*Loss*为损失函数（Loss Function）；*A*表示预测边界框的位置；*B*表示真实边界框的位置。

### 3.1.3 激活函数

YOLOX使用平滑的SiLU激活函数，它具有下界而无上界，如公式（3）所示。

$$\mathrm{SiLU}(x) = x \times \mathrm{Sigmoid}(x) = x \times \frac{1}{1 + e^{-x}} \quad (3)$$

### 3.1.4 无锚框方法

与以往的YOLO网络所使用的Anchor base不同，YOLOX采用Anchor free的策略，避免手动设置锚框和参数定义的步骤。根据无锚方法避免了手动锚设计和参数定义的步骤。无锚方法根据锚定框的中心点是否位于地面真值框的矩形内来确定正样本和负样本，可以有效简化检测器的训练和解码过程，缓解正负样本的不平衡，加快训练过程，并有助于一定程度的精度提高。

## 3.2 改进的YOLOX网络模型设计

YOLOX主干网络包含多个跨阶段层（Cross Stage Partial，CSP），具有多个残差网络，并且CSP层将输入特征经过少量处理与多个残差的输出特征直接相连，具有较大的残差边。该操作虽然有效缓解了加深网络带来的梯度消失问题，但这种残差会将特征信息连同包含的噪声一起传输到较深的网络，对主干网络提取特征造成一些不利的影响。因此，针对YOLOX在自然环境下火龙果检测的一些缺陷，本研究引入卷积注意力模块（Convolutional Block Attention Module，CBAM），在主干网络提取到的三个有效特征层后分别增加CBAM结构，使网络更加关注重要特征，抑制非必要特征。改进后的YOLOX网络结构如图2所示。

CBAM是Woo等[24]提出的一种注意力模块。CBAM会将输入的特征层分别进行通道注意力机制和空间注意力机制的处理，使提取到的特征具有更强的表示能力。其结构示意如图3所示。

### 3.2.1 通道注意力机制

通道注意力机制主要关注输入图像中有意义的信息。对输入尺寸为C×H×W的特征层分别进行平均池化和最大池化，获取到两个尺寸为C×1×1的特征层；接着它们经过两个神经网络MLP（Multilayer Perceptron），第一个神经元个数为C/r，激活函数为ReLU，第二个神经元个数为C。其中，C为输入的特征层通道数，r为下降率；将得到的两个结果进行相加，再经过一个Sigmoid函数，得到输入特征层每个通道的权重值；最后将获得的权重值乘以原输入特征层得到新的特征。对于一个输入特征，经过通道注意力机制后得到的特征如公式（4）所示。

$$\boldsymbol{F}' = \boldsymbol{M}_C(\boldsymbol{F}) \otimes \boldsymbol{F} \quad (4)$$

其中，$\boldsymbol{F}$表示输入的特征矩阵；$\boldsymbol{F}'$表示通道注意力机制输出的特征映射；$\boldsymbol{M}_c$为通道压缩权值矩阵；⊗为矩阵元素依次相乘。

### 3.2.2 空间注意力机制

空间注意力机制主要关注目标的位置信息。对经过通道注意力机制输出的特征层进行最大池化和平均池化，得到两个尺寸为1×H×W的特征层，将其进行堆叠；再经过一个通道数为1的7×7卷积操作和一个Sigmoid函数得到权重值；最后将权重值乘以输入的特征层得到最终的特征。对于通道注意力机制输出特征$\boldsymbol{F}'$，经过空间注意力机制后得到的特征$\boldsymbol{F}''$如公式（5）所示。

$$\boldsymbol{F}'' = \boldsymbol{M}_s(\boldsymbol{F}') \otimes \boldsymbol{F}' \quad (5)$$

其中，$\boldsymbol{F}''$表示空间注意力机制输出的特征矩阵；$\boldsymbol{M}_S$为空间压缩权值矩阵。

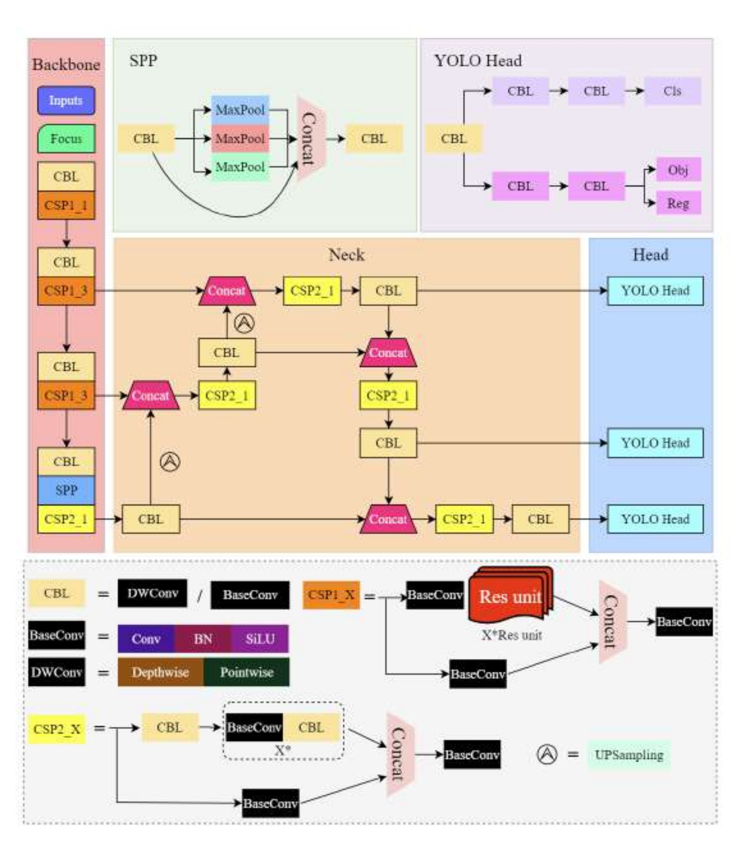



图 2　改进后的 YOLOX 网络结构

Fig. 2　The network structure of improved YOLOX

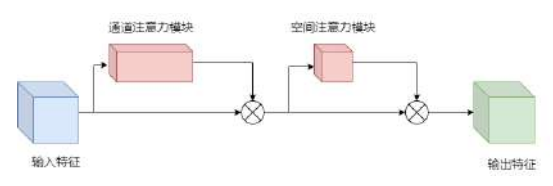



图 3　卷积注意力模块结构

Fig. 3　Convolutional Block Attention Module structure

## 3.3 模型的训练与测试

### 3.3.1　试验平台

本研究所使用的训练平台为配备 Windows10 64-bit 操作系统的台式计算机，其中 CPU 为 Intel i9 10900X @3.75GHz，GPU 为 NVIDIA GeForce GTX3090，RAM 为 128GB，PyTorch v1.7.1，CUDA v11.1，以及 CUDNN v8.0.2。训练环境和测试环境相同。

### 3.3.2　网络训练

将数据集中 5232 张图片按照 8:1:1 随机划分为训练集、验证集和测试集，具体如表 1 所示。使用随机梯度下降法（Stochastic Gradient Descent，SGD）和线性缩放（*lr*×BatchSize/64）的学习率，初始学习率 *lr*=0.001，余弦 *lr* 调度。权重衰减设置为 0.0005，SGD 的动量设置为 0.937。使用官方建议的预训练权重进行 300 轮次（Epoch）训练，其中前 50 轮次为冻结训练，批大小（BatchSize）设置为 64，之后的解冻训练将 BatchSize 设置为 32。最后，考虑到在 Mosaic 增强的过程中，样本由于偏离实际情况会导致训练效果降低。因此，最后 30 轮次关闭 Mosaic 和 MixUp 数据增强方式。

表 1　数据集的分布

Table 1　Distribution of the dataset

| 数据集 | 图片总数量/张 | 目标个数/个 |
|---|---|---|
| 训练集 | 4237 | 11,865 |
| 验证集 | 471 | 1340 |
| 测试集 | 524 | 1396 |
| 总数 | 5232 | 14,601 |

### 3.4 模型评价指标

使用调和均值 $F_1$ 值（F1-score）、召回率（Recall）、准确率（Precision）、平均精度（Average Precision，AP）、检测速度（Frames per Second，FPS）、平均检测时间（Average Time）、模型大小（Model Size）来评价训练后的模型。其中，$F_1$、Precision和Recall计算公式如下。

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP} \times 100\%} \tag{7}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN} \times 100\%} \tag{8}$$

其中，TP（True Positive，真阳性）指正确分割的正样本数量；FP（False Positive，假阳性）为未正确分割的正样本数量；FN（False Negative，假阴性）为未正确分割的负样本数量，$F_{1_{0.5}}$ 为在IoU=0.5的时候，模型的平均精度和召回率的调和平均值。

AP计算如公式（9）。

$$\text{AP} = \int_0^{1} dr\ \text{Precision} \times \text{Recall} \tag{9}$$

其中，$r$ 为积分变量，是对召回率与精确度乘积的积分；AP为PR（Precision-Recall）曲线与坐标轴围成的面积，取值在0～1之间；$AP_{0.5}$ 为IoU=0.5时不同查全率下的精度平均值；$AP_{0.5:0.95}$ 为将IoU的值从0.5到0.95，步长为0.5，计算这10个不同IoU情况下所有AP的平均值，可以更全面地反映检测模型的性能；$AP_S$、$AP_M$ 和 $AP_L$ 分别表示为IoU=0.5时，模型对小目标、中目标、大目标检测的平均精度，其中小目标定义为像素面积Area＜$32^2$，中目标 $32^2$＜Area＜$96^2$，大目标Area＞$96^2$。

## 4 结果与分析

### 4.1 不同分辨率输入的比较

对不同分辨率（640×640像素和416×416像素）火龙果图像的输入进行训练和比较。训练集和验证集的损失曲线如图4所示，红色和绿色曲线分别表示相同训练集下高分辨率和低分辨率输入的损失曲线。可见，在训练过程中，两种分辨率输入下模型的收敛性相似，在早期阶段训练集的损失迅速下降，随着训练轮次（Epoch）的增加，两条曲线逐渐下降并趋于稳定。当Epoch达到250左右时，模型逐渐收敛，最终红色曲线的损失值低于绿色曲线。

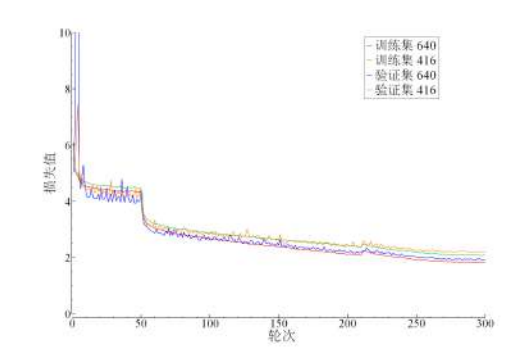



图4 不同输入分辨率的YOLOX-Nano损失曲线

Fig. 4 Loss curves of YOLOX-Nano with different input resolutions

为了分析模型在验证集下的收敛情况，在相同验证集下，蓝色和黄色曲线分别表示高分辨率和低分辨率输入的验证集损失曲线，两条曲线收敛性相似，最终达到平缓状态，蓝色曲线的损失值低于黄色曲线，与训练集下两者损失情况相似。

综上表明，两种不同分辨率输入下的YOLOX-Nano网络都具有较好的收敛性，而高分辨率输入使火龙果特征更加清晰，可以更好地学习特征，该结论与文献类似[25]。因此，本研究使用640×640像素输入分辨率作为最终检测模型。

### 4.2 不同模型大小的结果分析

使用相同的训练方法，分别训练了三种轻量化的YOLOX网络模型（即YOLOX-Nano、YOLOX-Tiny和YOLOX-S），并在相同测试集下进行测试，分析其性能差异。由表2可知，三种检测网络 $F_1$ 值相差不大，在IoU的值为0.5的时候，YOLOX-Nano的平均精度相较YOLOX-Tiny和YOLOX-S分别提高了2.3个百分点和1.6个百分

点。而YOLOX-Nano的$AP_{0.5:0.95}$值达到70.2%，均高于YOLOX-Tiny和YOLOX-S的$AP_{0.5:0.95}$值3.9个百分点和2.6个百分点。YOLOX-Nano模型大小仅3.7 MB，远小于YOLOX-Tiny的19.4 MB和YOLOX-S的34.3 MB。上述结果中，模型大小最小的YOLOX-Nano性能表现最优，这可能是由于在训练过程中，三种模型保持几乎相同的学习进度和优化参数，训练中对三种模型的增强策略相同（即消除Mosic和MaxUp的时期相同）。而合适的增强策略在不同大小的模型中有所不同，对于大模型，更强的增强更有用，文献［19］中有类似的结论。因此，为了便于在嵌入式和移动设备上部署，在保证有较好的检测精度的前提下，本研究选择模型较小的YOLOX-Nano作为基准进行研究。

**表2 不同模型大小的YOLOX-Nano检测结果对比**

Table 2 Comparison of YOLOX test results of different model sizes

| 模型 | 大小/MB | $F_{1_{0.5}}$ | $AP_{0.5:0.95}$/% | $AP_{0.5}$/% |
|---|---|---|---|---|
| YOLOX-Nano | 3.7 | 0.98 | 70.2 | 98.0 |
| YOLOX-Tiny | 19.4 | 0.97 | 66.3 | 95.7 |
| YOLOX-S | 34.3 | 0.97 | 67.6 | 96.4 |

## 4.3 改进后YOLO网络模型的结果分析

为验证融合了CBAM后的YOLOX-Nano网络模型性能，对改进前后的网络模型进行比较分析。分别将两个模型在相同火龙果数据集上进行识别试验。如图5所示，改进前的YOLOX-Nano在逆光条件下受太阳光照影响，存在误识别现象，错误地将逆光画面中产生的一定几何外形的光斑识别为火龙果，而改进后的YOLOX-Nano模型则未出现这种情况。在图6中，两种网络模型均对场景中的火龙果进行正确识别，YOLOX-Nano识别到的目标从左到右置信度分别为0.73、0.79和0.69，而改进后的YOLOX-Nano识别到的相同果实置信度从左到右分别为0.82、0.81和0.74，引入CBAM后的YOLOX-Nano识别到火龙果的置信度高于YOLOX-Nano。两种模型检测火龙果的具体评价指标结果如表3所示。



（a）原始YOLOX-Nano



（b）改进的YOLOX-Nano

图5 YOLOX-Nano改进前后火龙果检测结果对比

Fig. 5 Comparison of test results of YOLOX-Nano before and after improvement



（a）原始YOLOX-Nano，置信度由左到右分别为0.73，0.79和0.69



（b）改进的YOLOX-Nano，置信度由左到右分别为0.82，0.81和0.74

图6 YOLOX-Nano改进前后火龙果检测结果对比

Fig. 6 Test results comparison of YOLOX-Nano before and after the improvement

表3 改进的YOLOX-Nano网络火龙果检测结果对比

Table 3 Comparison of dragon fruit detection results of improved YOLOX-Nano network

| 改进的网络模型 | 模型大小/MB | 平均时间/ms | $F_{1_{0.5}}$ | $AP_{0.5}$/% | $AP_{0.5:0.95}$/% | $AP_S$/% | $AP_M$/% | $AP_L$/% |
|---|---|---|---|---|---|---|---|---|
| YOLOX-Nano | 3.70 | 18.46 | 0.98 | 98.0 | 70.2 | 55.4 | 60.8 | 76.2 |
| YOLOX-Nano+CBAM | 3.76 | 21.72 | 0.99 | 98.9 | 72.4 | 56.2 | 63.3 | 78.0 |

由表3可知，本研究提出的改进后的YOLOX-Nano目标检测网络，在检测时间相差3.26 ms的情况下，$F_1$值达到0.99，$AP_{0.5}$值高于YOLOX-Nano目标检测网络0.9个百分点，$AP_{0.5:0.95}$值提升了2.2个百分点。在添加了CBAM后，模型大小略有增加，但模型对大、中、小不同尺度的火龙果目标检测平均精度分别提升1.8%、2.5%和0.8%。

对比结果说明，在复杂自然背景下，引入CBAM一定程度上提高了模型对火龙果目标检测的鲁棒性。通道注意力机制对特征图中不同通道赋予不同的权重值，空间注意力机制对同一特征图不同位置的特征点赋予不同权重值，二者的结合使得提取到的火龙果特征更加精炼，总体上提升了模型检测的平均精度，其模型体积和检测时间满足嵌入式设备部署需求。

## 4.4 不同YOLO网络模型的结果分析

为客观地反映改进的YOLOX-Nano网络的性能，本研究还使用相同的设置训练其它轻量化YOLO模型以及标准的YOLOv3模型作为参照，并将它们与C-YOLOX-Nano模型进行比较，五种目标检测网络检测对比结果如表4所示。

表4 不同网络火龙果检测结果对比

Table 4 Comparison of dragon fruit detection results of different networks

| 网络模型 | 模型大小/MB | 帧率/(f·s$^{-1}$) | 平均时间/ms | $F_{1_{0.5}}$ | $AP_{0.5}$/% | $AP_{0.5:0.95}$/% | $AP_S$/% | $AP_M$/% | $AP_L$/% |
|---|---|---|---|---|---|---|---|---|---|
| YOLOX-Nano+CBAM | 3.76 | 46 | 21.72 | 0.99 | 98.9 | 72.4 | 56.2 | 63.3 | 78.0 |
| YOLOv5-S | 27.10 | 59 | 16.87 | 0.93 | 91.0 | 59.5 | 29.0 | 49.8 | 66.5 |
| MobileNetV3-YOLOv4 | 53.70 | 44 | 22.66 | 0.94 | 91.8 | 55.4 | 38.8 | 44.8 | 62.6 |
| YOLOv4-Tiny | 22.40 | 145 | 6.88 | 0.91 | 89.1 | 54.4 | 25.5 | 42.6 | 62.2 |
| YOLOv3 | 235.00 | 51 | 19.38 | 0.83 | 72.7 | 41.4 | 1.7 | 31.0 | 49.7 |

可以看出，在相同测试集下改进后的YOLOX-Nano网络的$AP_{0.5}$值达到了98.9%，均高于其它模型。此外，该网络的$AP_{0.5:0.95}$值分别比YOLOv5-S、MobileNetV3-YOLOv4和YOLOv4-Tiny提升了12.9%、17.0%和18.0%。观察$AP_L$、$AP_M$、$AP_S$的值不难看出，在近距离视野范围内的大目标火龙果检测上，改进的YOLOX-Nano网络模型对比其它轻量化YOLO网络模型的平均检测精度提升了约12%左右。随着距离的增加，对中远距离视野范围内火龙果的像素面积逐渐减少，检测精度均有所下降，改进的YOLOX-Nano对中目标火龙果的检测性能对比其它轻量化模型提升了约17%左右。对小目标火龙果检测的平均精度值为56.2%，大约是YOLOv5-S的1.9倍，是MobileNetV3-YOLOv4的1.4倍，是YOLOv4-Tiny的2.1倍。而原始的YOLOv3模型在检测精度上均小于其它模型，对小目标火龙果的$AP_S$值仅为1.7%，基本无法对远距离火龙果进行检测。

本研究提出的改进的YOLOX-Nano网络模型在保持模型大小更轻量化的同时，在检测性能上显著优于其它模型，$F_1$值达到0.99，远高于其它模型。虽然检测速度不是最快的，但改进的YOLOX-Nano仍然具有较高的检测速度，其帧率达到46 f/s，平均检测时间约21.72 ms，可以保证火龙果采摘机器人检测的实时性。

为了验证改进后的YOLOX-Nano在自然环境中的检测效果，本研究将不同模型对顺光、遮阴、逆光等不同自然环境下的火龙果检测效果进行对比。对于图7和图8所示的顺光和遮阴条件下，显然可以看出，本研究中改进的YOLOX-

Nano 网络检测结果均好于其它 4 种目标检测网络，并未出现误检测情况，甚至对于较远处的小目标火龙果和表面遮挡严重的火龙果也被识别到。对于图 9 所示的严重逆光情况下，上述网络均存在漏检的情况，仅有改进的 YOLOX-Nano 和 YOLOV3 网络检测出了较近的火龙果，且改进的 YOLX-Nano 检测到的置信度略高于 YOLOv3，而其它 3 种网络在该自然情况下均未检测出火龙果。

综上所述，本研究提出的方法能够有效地对自然环境中火龙果进行检测。此外，在严重逆光情况下，会影响算法的检测性能，解决此类环境光照对相机影响的问题仍然需要进一步的研究。




(a)YOLOX-Nano+CBAM


(b)YOLOv5-S


(c)MobileNetV3-YOLOv4


(d)YOLOv4-Tiny


(e)YOLOv3

图 7 顺光时不同网络火龙果检测效果

Fig. 7 Dragon fruit detection effect of different networks during lighting


(a)YOLOX-Nano+CBAM


(b)YOLOv5-S


(c)MobileNetV3-YOLOv4


(d)YOLOv4-Tiny


(e)YOLOv3

图 8 遮阴时不同网络火龙果检测效果

Fig. 8 Dragon fruit detection effect of different networks during shading



(a)改进的 YOLOX-Nano



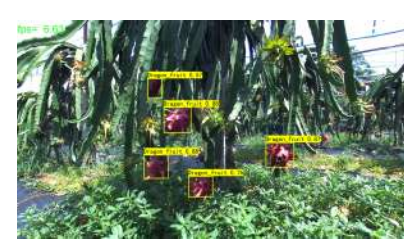
(b)YOLOv3

图 9 逆光时不同网络火龙果检测效果

Fig. 9 Dragon fruit detection effect of different networks during backlighting

## 4.5 不同视频分辨率输入的结果分析

为了验证改进后的 YOLOX-Nano 目标检测网络是否满足自然环境下实时检测的需求，本研究使用 ZED 2i 双目相机采集不同分辨率输入下的视频数据进行检测，计算改进后的 YOLOX-Nano 网络在各分辨率下 0～100 帧的平均检测帧率，对比结果如表 5 所示，其中分辨率为相机采样分辨率，相机帧率为相机采样设置帧率，平均检测帧率为改进后的 YOLOX-Nano 网络检测帧率的平均值。该模型在 4416×1242 分辨率下平均检测帧率为 5.77 f/s，降低输入分辨率，在 1340×376 分辨率下平均检测帧率提高到 20.94 f/s，随着分辨率输入的降低，模型平均检测帧率提升。图 10 为改进后的 YOLOX-Nano 网络在 3840×1080 和 1340×376 分辨率下左相机视角下第 50 帧的检测效果图，此时检测帧率分别为 6.63 f/s 和 20.41 f/s。综上，改进后的 YOLOX-Nano 目标检测网络检测帧率满足实时检测需求，可为其他自然环境下水果识别提供一定的参考。



(a)3840×1080 分辨率(ZED 2i 左相机视图)



(b)1340×376 分辨率(ZED 2i 左相机视图)

图 10 改进的 YOLOX-Nano 网络火龙果检测帧率对比

Fig. 10 Comparison of dragon fruit detection frame rates of improved YOLOX-Nano network

**表 5 不同分辨率视频火龙果检测结果**

Table 5 Comparison of dragon fruit detection frame rates in different resolutions

| 分辨率 | 相机帧率/(f·s$^{-1}$) | 平均检测帧率/(f·s$^{-1}$) |
|---|---|---|
| 4416×1242 | 15 | 5.77 |
| 3840×1080 | 30 | 6.89 |
| 2560×720 | 60 | 12.34 |
| 1340×376 | 100 | 20.94 |

## 5 结 论

本研究介绍了一种基于改进的 YOLOX 火龙果检测方法。通过引入注意力机制改进 YOLOX-Nano 目标检测网络。网络经过训练后，在 IoU=0.5 时检测火龙果的 AP 值为 98.9%。在 $AP_{0.5:0.95}$ 的值为 72.4%，对小目标检测 $AP_S$ 为 56.2%，此外，改进后的网络模型与其他网络模型检测方法的比较，该模型在不同光照和遮挡条件下准确地检测出火龙果。改进的 YOLOX-Nano 单幅图像的平均检测时间为 21.72 ms，且模型大小为 3.76 MB 的轻量级模型，适合用于嵌入式设备和移动终端。对不同输入分辨率的视频流进行实时检测，当输入分辨率为 1340×376 时，平均检测帧率约为 21 f/s，满足实时检测需求。综上，本研究提出改进的 YOLOX-Nano 目标检测模型的实时性

和准确性满足在自然环境中火龙果的快速检测要求，对火龙果智能化收获装备的研制具有重要意义，可为其他水果的智能化检测技术研究提供参考。

**参考文献：**


[1] 曾曦, 胡桂兵, 秦永华. 广东省火龙果产业发展现状及对策[J]. 中国果业信息, 2019, 36(9): 9-12.

[2] 徐磊磊, 金琰, 侯媛媛, 等. 我国火龙果市场与产业调查分析报告[J]. 农产品市场, 2021(8): 43-45.

[3] 陆华忠, 李君, 李灿. 果园机械化生产技术研究进展[J]. 广东农业科学, 2020, 47(11): 226-235.
LU H,LI J,LI C. Research progress in orchard mechanization production technology[J]. Guangdong Agricultural Sciences, 2020, 47(11): 226-235.

[4] 郑太雄, 江明哲, 冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. 仪器仪表学报, 2021, 42(9): 28-51.
ZHENG T, JIANG M, FENG M. Vision based target recognition and location for picking robot: A review[J]. Chinese Journal of Scientific Instrument, 2021, 42(9): 28-51.

[5] JORDI G-M, RICARDO S-C, JOAN R R-P, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry[J]. Computers and Electronics in Agriculture, 2020, 169: ID 105165.

[6] 李志军, 杨圣慧, 史德帅, 等. 基于轻量化改进YOLOv5的苹果树产量测定方法[J]. 智慧农业(中英文), 2021, 3(2): 100-114.
LI Z, YANG S, SHI D, et al. Yield estimation method of apple tree based on improved lightweight YOLOv5[J]. Smart Agriculture, 2021, 3(2): 100-114.

[7] JI J, XUZ, MA H, et al. Apple fruit recognition based on a deep learning algorithm using an improved lightweight network[J]. Applied Engineering in Agriculture, 2021, 37(1): 123-134.

[8] 穆龙涛, 高宗斌, 崔永杰, 等. 基于改进AlexNet的广域复杂环境下遮挡猕猴桃目标识别[J]. 农业机械学报, 2019, 50(10): 24-34.
MU L, GAO Z, CUI Y, et al. Kiwifruit detection of far-view and occluded fruit based on improved AlexNet[J]. Transactions of the CSAM, 2019, 50(10): 24-34.

[9] 陈燕, 王佳盛, 曾泽钦, 等. 大视场下荔枝采摘机器人的视觉预定位方法[J]. 农业工程学报, 2019, 35(23): 48-54.
CHEN Y, WANG J, ZENG Z, et al. Vision pre-positioning method for litchi picking robot under large field of view[J]. Transactions of the CSAE, 2019, 35(23): 48-54.

[10] 彭红星, 李荆, 徐慧明, 等. 基于多重特征增强与特征融合SSD的荔枝检测[J]. 农业工程学报, 2022, 38(4): 169-177.
PENG H, LI J, XU H, et al. Litchi detection based on multiple feature enhancement and feature fusion SSD[J]. Transactions of the CSAE, 2022, 38(4): 169-177.

[11] ROY ARUNABHA M, JAYABRATA BHADURI. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4[J]. Computers and Electronics in Agriculture, 2022, 193: ID 106694.

[12] XU Z, JIA R, SUN H, et al. Light-YOLOv3: Fast method for detecting green mangoes in complex scenes using picking robots[J]. Applied Intelligence, 2020, 50(12): 4670-4687.

[13] XU P, FANG N, LIU N, et al. Visual recognition of cherry tomatoes in plant factory based on improved deep instance segmentation[J]. Computers and Electronics in Agriculture, 2022, 197: ID 106991.

[14] 周云成, 许童羽, 郑伟, 等. 基于深度卷积神经网络的番茄主要器官分类识别方法[J]. 农业工程学报, 2017, 3(15): 219-226.
ZHOU Y, XU T, ZHENG W, et al. Classification and recognition approaches of tomato main organs based on DCNN[J]. Transactions of the CSAE, 2017, 33(15): 219-226.

[15] ZHENG T, JIANG M, LI Y, et al. Research on tomato detection in natural environment based on RC-YOLOv4[J]. Computers and Electronics in Agriculture, 2022, 198: ID 107029.

[16] 赵春江, 文朝武, 林森, 等. 基于级联卷积神经网络的番茄花期识别检测方法[J]. 农业工程学报, 2020, 36(24): 143-152.
ZHAO C, WEN C, LIN S, et al. Tomato florescence recognition and detection method based on cascaded neural network[J]. Transactions of the CSAE, 2020, 36(24): 143-152.

[17] LI X, QIN Y, WANG F, et al. Pitaya detection in orchards using the MobileNet-YOLO model, 2020[C]// Technical Committee on Control Theory. Beijing, China: Chinese Association of Automation, 2020.

[18] 王金鹏, 高凯, 姜洪喆, 等. 基于改进的轻量化卷积神经网络火龙果检测方法(英文)[J]. 农业工程学报, 2020, 36(20): 218-225.
WANG J, GAO K, JIANG H, et al. Method for detecting dragon fruit based on improved lightweight convolutional neural network[J]. Transactions of the CSAE, 2020, 36(20): 218-225.

[19] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J/OL]. arXiv: 2107.08430, 2021.

[20] ZHANG J, KE S. Improved YOLOX fire scenario detection method[J]. Wireless Communications and Mobile Computing, 2022, 2022: 1-8.

[21] LIU B, HUANG J, LIN S, et al. Improved YOLOX-S abnormal condition detection for power transmission line corridors, 2021[C]// 2021 IEEE 3rd International Conference on Power Data Science (ICPDS). Piscataway, New York, USA: IEEE, 2021: 13-16..
[22] LIU M, ZHU C. Residual YOLOX-based Ship Object Detection Method, 2022[C]// 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). Piscataway, New York, USA: IEEE, 2022: 427-431.
[23] WANG J, TANG C, LI J. Towards real-time analysis of marine phytoplankton images sampled at high frame rate by a YOLOX- based object detection algorithm[C]// OCEANS 2022-Chennai. Piscataway, New York, USA: IEEE, 2022: 1-9.
[24] WOO S, PARK L, LEE J Y, et al. CBAM: convolutional block attention module[J/OL]. arXiv: 1807. 06521, 2018.
[25] FU L, WU F, ZOU X, et al. Fast detection of banana bunches and stalks in the natural environment based on deep learning[J]. Computers and Electronics in Agriculture, 2022, 194: ID 106800.

# Detection Method for Dragon Fruit in Natural Environment Based on Improved YOLOX

SHANG Fengnan[1, 2, 3], ZHOU Xuecheng[1, 2, 3*], LIANG Yingkai[1, 2, 3], XIAO Mingwei[1, 2, 3], CHEN Qiao[1, 2, 3], LUO Chendi[1, 2, 3]

(1. *College of Engineering*, *South China Agricultural University*, *Guangzhou* 510642, *China*; 2. *Guangdong Provincial Key Laboratory of Agricultural Artificial Intelligence*, *Guangzhou* 510642, *China*; 3. *Key Laboratory of Key Technology on Agricultural Machine and Equipment*, *Ministry of Education*, *Guangzhou* 510642, *China*)

**Abstract:** Dragon fruit detection in natural environment is the prerequisite for fruit harvesting robots to perform harvesting. In order to improve the harvesting efficiency, by improving YOLOX (You Only Look Once X) network, a target detection network with an attention module was proposed in this research. As the benchmark, YOLOX-Nano network was chose to facilitate deployment on embedded devices, and the convolutional block attention module (CBAM) was added to the backbone feature extraction network of YOLOX-Nano, which improved the robustness of the model to dragon fruit target detection to a certain extent. The correlation of features between different channels was learned by weight allocation coefficients of features of different scales, which were extracted for the backbone network. Moreover, the transmission of deep information of network structure was strengthened, which aimed at reducing the interference of dragon fruit recognition in the natural environment as well as improving the accuracy and speed of detection significantly. The performance evaluation and comparison test of the method were carried out. The results showed that, after training, the dragon fruit target detection network got an $AP_{0.5}$ value of 98.9% in the test set, an $AP_{0.5:0.95}$ value of 72.4% and $F_1$ score was 0.99. Compared with other YOLO network models under the same experimental conditions, on the one hand, the improved YOLOX-Nano network model proposed in this research was more lightweight, on the other hand, the detection accuracy of this method surpassed that of YOLOv3, YOLOv4 and YOLOv5 respectively. The average detection accuracy of the improved YOLOX-Nano target detection network was the highest, reaching 98.9%, 26.2% higher than YOLOv3, 9.8% points higher than YOLOv4-Tiny, and 7.9% points higher than YOLOv5-S. Finally, real-time tests were performed on videos with different input resolutions. The improved YOLOX-Nano target detection network proposed in this research had an average detection time of 21.72 ms for a single image. In terms of the size of the network model was only 3.76 MB, which was convenient for deployment on embedded devices. In conclusion, not only did the improved YOLOX-Nano target detection network model accurately detect dragon fruit under different lighting and occlusion conditions, but the detection speed and detection accuracy showed in this research could able to meet the requirements of dragon fruit harvesting in natural environment requirements at the same time, which could provide some guidance for the design of the dragon fruit harvesting robot.

**Key words:** fruits picking; natural environment; dragon fruit; object detection; YOLOX; attention mechanism; deep learning